Complementary Ionization Techniques for the Analysis of Scotch Whisky by High Resolution Mass Spectrometry

Supporting Information

Complementary Ionisation Techniques for the Analysis of Scotch Whisky as a Complex Mixture by High Resolution Mass Spectrometry

Will Kew[1], C. Logan Mackay[1], Ian Goodall[2], David J. Clarke[1*], Dušan Uhrín[1*]

[1] EaStCHEM School of Chemistry, Joseph Black Building, University of Edinburgh, Edinburgh, UK, EH9 3FJ.

[2] The Scotch Whisky Research Institute, The Robertson Trust Building, Research Avenue North, Riccarton, Edinburgh, UK, EH14 4AP.

Corresponding Authors:

* Prof Dušan Uhrín, EaStCHEM, School of Chemistry, Joseph Black Building, University of Edinburgh, Edinburgh, UK, EH9 3FJ, Tel: +44(0)131 650 4742, email: dusan.uhrin@ed.ac.uk
* Dr David Clarke, EaStCHEM, School of Chemistry, Joseph Black Building, University of Edinburgh, Edinburgh, UK, EH9 3FJ, Tel: +44(0)131 650 4808, email: dave.clarke@ed.ac.uk

Contents

Positive Mode Data Acquisition

Positive mode data requires higher resolving powers for unambiguous and confident assignment. This is because in positive mode, formula adducts are more likely. In provisional studies, we observed protonated species in positive modes in all four ionisation techniques (APCI, APPI, ESI, and LDI), as well as sodium adducts in ESI and potassium adducts in LDI. All four methods had many peaks, and complete unambiguous formula assignment was not possible on our instrument.

Calculations were performed to determine what possible mass differences could occur in complex mixtures of CHO species. All possible, chemically logical, monoisotopic formulae were generated using scripts published previously.[1] Possible formulae at 401 *m/z* and 799 *m/z* in both positive and negative mode were investigated as an example. Negative formulae were generated as deprotonated ions, positive species were protonated, or as sodium or potassium adducts. Resolving power calculations were based on a 12 Tesla magnet.

$$Measured\ Resolving\ Power = \frac{m}{\Delta m_{50\%}} \qquad \text{Eq. 1}$$

$$Theoretical\ Resolving\ Power = \frac{CBT}{m/z} \qquad \text{Eq. 2}$$

The resolving power is defined as the mass of a peak divided by its full width at half maximum ($\Delta m_{50\%}$) as shown in Equation 1. The theoretical resolving power can be calculated from the magnetic field strength (B, in Tesla), the length of the FID (T, in seconds), the mass of interest (m/z) and a constant (C) as shown in Equation 2.[2] For our data acquired on our instrument, C was determined to be approximately $1x10^7$.

Results for these calculations are summarised in Table S1. Here we have calculated resolving power required based on the peaks being of equal intensity and thus not requiring baseline resolution. For a large difference in ion intensities, higher resolving powers would be needed. All calculations have been made for a magnetic field of 12T and 4 MW data, corresponding to an acquisition length of 1.1185 s, processed in magnitude mode with zero filling once and full sine apodization.

It is clear that as mass increases the number of chemically logical, monoisotopic formula increases as well. The relative (ppm) gaps between species decreases as mass increases. The absolute, mDa, smallest gap does not change between the two nominal masses in the positive mode as the smallest gap is due to a simple $-C_3Na + OH_4K$ change. If both sodium and potassium adducted ions are possible, it is impossible to discriminate between these assignments without a very high resolving power. In negative mode, the smallest gap is much larger, as it is due to much more significant compositional changes. It may be possible to assign formula based on other criteria than just mass accuracy, for example the types of compounds expected in a sample, or classes of compounds to be ionised by a given source – i.e. aromatics in APPI, fatty acids in ESI.

Table S1 - Theoretical complexity of CHO spectra in positive and negative ionisation modes, showing minimum distances between peaks and the required resolving power and transient length at 12T required to resolve them. Smallest gaps are shown in both mDa and (parts-per-million).

Polarity	Mass (*m/z*)	# Possible CHO Compounds	Smallest Gap mDa (ppm)	Formula Change	Resolving Power Required	Transient Length (s)
Positive	401	164	0.152 (0.38)	$-C_3 +H_4 +O -Na +K$	2,638,223	8.8
Positive	799	621	0.152 (0.19)	$-C_3 +H_4 +O -Na +K$	5,257,105	35
Negative	401	60	3.5 (8.8)	$+C_{18} -H_8 -O_{13}$	115,000	0.38
Negative	799	211	2.37 (2.97)	$-C_{25} +H_{12} +O_{18}$	337,168	2.2

These results confirm that without higher resolution data, most probably requiring a higher resolution instrument, it is impossible to confidently assign positive mode data of complex mixtures from broadband mass spectra. By use of phased, absorption mode spectra, the resolution could be increased between √2 and 2. With 2ω quadrupolar detection, the resolution could be doubled again (or transient length halved).[3–5] At a higher magnetic field, resolution would also increase. However, the improvement at 21T compared to 12T is only 1.75 x (=21/12, Eq. 2), and there currently are only two 21T FTICR instruments in the world. Combination of these three improvements results in a net ~5.6-fold improvement in resolution for the same length of transient. This alone is obviously still insufficient for short transients, but longer transients (6 seconds or longer) are possible with improved ICR cell design and control, vacuum levels, and ion counts.

These calculations were made for a relatively simple case of CHO only formula singly charged as deprotonated (negative) or protonated (positive) or sodium or potassium adducts (positive). Other ionisation states were not considered. Radical ions, for example, would appear ±1 *m/z* relative to their (de)protonated ion counterpart, and are therefore not a source of ambiguity for monoisotopic formula assignment here. Addition of more heteroatoms, e.g. NSP, would increase difficulty further again. Also, these calculations are for monoisotopic formulae only; inclusion of isotopologue peaks would further increase the difficulty of the assignment. Finally, these calculations are based on ideal spectra of equal ion counts. Large differences in ion counts, unstable ion-clouds, space charge effects, peak coalescence, and thermal noise will aggravate the situation. Of course, no sample is going to contain every possible formula. Overall, though, it is easy to see that negative mode data with simply deprotonated formula are much more straightforward to acquire with sufficient resolution, and to assign, than positive mode data.

From a practical perspective, positive mode data were acquired without any substantial complications compared to the negative mode data. If one could be confident of only producing one type of ion (e.g. all protonated, no adducts), it may be possible to assign data at our resolution, as the minimum gaps would be comparable to those for negative mode (where only one type of ion is expected).

Calibration Lists

Table S2 - Calibration lists for APCI, APPI, ESI and LDI in negative mode containing molecular ion formulae and exact masses. All formulae are in $[M-H]^-$ form

APCI (-)		APPI (-)		ESI (-)		LDI (-)	
Formula	***m/z***	**Formula**	***m/z***	**Formula**	***m/z***	**Formula**	***m/z***
C8H7O3	151.040068	C6H9O5	161.046644	C4H5O6	149.009161	C10H7O3	175.040068
C6H9O5	161.045547	C7H5O5	169.014247	C9H17O2	157.123403	C11H9O3	189.055718
C8H7O4	167.034982	C6H11O6	179.056111	C7H5O5	169.014247	C12H11O3	203.071368
C10H19O2	171.139053	C6H11O7	195.051026	C10H19O2	171.139053	C13H13O3	217.087018
C9H9O4	181.050632	C8H11O7	219.051026	C6H9O6	177.040462	C12H7O5	231.029897
C10H19O3	187.133968	C8H13O8	237.061591	C6H11O7	195.051026	C13H9O5	245.045547
C12H23O2	199.170353	C9H15O8	251.077241	C12H23O2	199.170354	C14H11O5	259.061197
C11H11O4	207.066282	C9H15O9	267.072155	C11H11O4	207.066282	C15H13O5	273.076847
C12H11O4	219.066282	C10H17O9	281.087806	C8H13O7	221.066676	C16H15O5	287.092497
C11H13O5	225.076847	C11H17O9	293.087806	C8H13O8	237.061591	C14H5O8	300.998991
C12H9O5	233.045547	C14H5O8	300.998991	C9H13O8	249.061591	C17H17O5	301.108147
C15H29O2	241.217304	C12H17O9	305.087806	C12H5O7	261.004076	C18H19O5	315.123797
C13H13O5	249.076847	C11H19O10	311.098370	C11H15O8	275.077241	C19H21O5	329.139447
C16H29O2	253.217304	C12H19O10	323.098370	C12H15O8	287.077241	C20H23O5	343.155097
C16H31O2	255.232954	C13H15O10	331.067070	C13H7O8	291.014641	C18H13O8	357.061591
C13H13O6	265.071762	C12H21O11	341.108935	C13H15O8	299.077241	C19H15O8	371.077241
C15H13O5	273.076847	C16H17O9	353.087806	C14H5O8	300.998991	C20H17O8	385.092891
C18H35O2	283.264254	C17H19O9	367.103456	C13H15O9	315.072156	C21H19O8	399.108541
C17H23O4	291.160183	C18H17O9	377.087806	C13H15O10	331.067070	C22H21O8	413.124191
C17H15O5	299.092497	C19H21O9	393.119106	C12H21O11	341.108935	C23H23O8	427.139841
C14H5O8	300.998991	C20H21O9	405.119106	C15H19O9	343.103456	C23H21O9	441.119106
C16H15O6	303.087412	C15H25O13	413.130064	C16H19O9	355.103456	C24H23O9	455.134756
C12H21O9	309.119106	C15H25O14	429.124979	C17H21O9	369.119106	C25H25O9	469.150406
C17H15O6	315.087412	C15H27O15	447.135544	C19H17O9	389.087806	C26H27O9	483.166056
C18H15O6	327.087412	C16H27O15	459.135544	C20H19O9	403.103456	C26H25O10	497.145321
C19H17O6	341.103062	C17H27O15	471.135544	C20H21O10	421.114020	C27H27O10	511.160971
C18H15O7	343.082326	C18H29O15	485.151194	C20H23O11	439.124585	C28H29O10	525.176621
C21H19O6	367.118712	C18H29O16	501.146108	C19H25O13	461.130064	C29H31O10	539.192271
C24H47O2	367.358154	C30H45O7	517.317077	C17H27O15	471.135544	C30H33O10	553.207921
C19H21O8	377.124191	C18H31O17	519.156673	C22H25O12	481.135150	C31H35O10	567.223571
C24H45O4	397.332333	C19H33O17	533.172323	C22H25O13	497.130064	C30H29O12	581.166450
C21H19O8	399.108541	C20H33O17	545.172323	C18H31O16	503.161758	C31H31O12	595.182100
C22H19O8	411.108541	C21H33O17	557.172323	C24H27O12	507.150800	C32H33O12	609.197750
C25H47O4	411.347983	C20H35O18	563.182888	C30H45O7	517.317077	C33H35O12	623.213400
C22H21O8	413.124191	C22H33O17	569.172323	C30H45O8	533.311992	C34H37O12	637.229050
C22H23O9	431.134756	C21H33O18	573.167238	C24H29O14	541.156279	C36H27O12	651.150800
C22H25O10	449.145320	C20H35O19	579.177802	C25H27O14	551.140629	C37H29O12	665.166450
C24H15O10	463.067070	C22H35O18	587.182888	C29H31O12	571.182100	C38H31O12	679.182100
C26H25O9	481.150406	C21H35O19	591.177802	C28H37O13	581.223965	C39H33O12	693.197750
C29H39O6	483.275212	C22H37O19	605.193452	C22H35O19	603.177802	C40H35O12	707.213400

C25H17O11	493.077635	C21H37O20	609.188367	C30H47O11S	615.284457	C41H37O12	721.229050
C25H17O12	509.072549	C24H37O19	629.193452	C24H39O19	631.209102		
C30H45O7	517.317077	C24H37O20	645.188367	C27H23O18	635.088987		
C28H29O10	525.176621	C24H39O20	647.204017	C36H55O12	679.369901		
C28H29O11	541.171535	C36H57O12	681.385551	C36H57O13	697.380465		
C29H29O11	553.171535	C37H51O12	687.338600	C27H43O23	735.220061		
C30H31O11	567.187185	C36H57O13	697.380465	C44H51O15	819.323345		
C29H33O12	573.197750			C42H67O18	859.433289		
C30H33O12	585.197750			C45H33O29	1037.111299		
C32H33O11	593.202835			C47H37O29	1065.142599		
C32H31O12	607.182100						
C32H35O13	627.208315						
C33H33O13	637.192665						
C33H39O14	659.234529						
C33H41O15	677.245094						
C38H39O13	703.239615						
C39H39O13	715.239615						
C38H41O14	721.250179						
C39H39O14	731.234529						
C39H39O15	747.229444						
C41H43O15	775.260744						

Assignment of Formulas

Formula were assigned using the software 'Formularity'.[6,7]

Formularity settings were as follows:

- Alignment was enabled
    - 1.0 ppm tolerance
- The CIA database was 'CIA_DB_2016_11_21.bin'.
    - Formula tolerance was 0.25 ppm
    - The DB mass limit was 700.000
    - The formula score was 'The lowest error'
- Formula filters were used:
    - Special Filter – "None:"
    - User-defined filter: "C>0 AND N=0 AND S=0 AND P=0"
- Relationships were enabled.
    - Max gaps: 5
    - Error: 1.0 ppm
    - Formula building blocks were:
        - CH2
        - H2
        - CO2
        - O
- For deprotonated negative ion search:
    - Charge: -1
    - Ionization: proton_detachment
    - Result: "M-p"
- For radical anion search:
    - Charge: -1
    - Ionization: electron_attachment
    - Result: "M+e"

All 16 spectra (4 samples, 4 ionisation sources) were assigned simultaneously, with peak alignment prior to formula assignment.

Combination of the deprotonated assignment data and radical assignment data was performed with a Python script. There were no conflicts of assignments between radical and deprotonated forms.

Table S3 – Summary statistics for each spectrum – each sample in each ionisation mode. Numbers shown include total number of peaks, monoisotopic assignments, percentage of which were radical anions, total percentage of formula assigned, and the mean error of assignment in parts-per-billion. This table expands on the detail provided in Table 1 in the main paper.

Mode	Total # Peaks	# of Monoisotopic Assignments	Percentage Monoisotopic Radical Anion	Percentage Formula Assigned	Mean Error of Assignment (ppb)
APCI*	**2795**	**2070**	**30.9%**	**89.9%**	**57**
S14-1941	3980	2805	34.7%	84.5%	67
S14-1944	1549	1342	27.9%	96.0%	46
S14-1962	4347	3000	36.0%	83.4%	68
S14-2196	1305	1131	24.8%	95.9%	46
APPI*	**895**	**777**	**10.2%**	**95.3%**	**40**
S14-1941	1295	1080	18.4%	93.7%	48
S14-1944	531	499	2.4%	96.8%	33
S14-1962	717	654	6.0%	96.5%	36
S14-2196	1038	875	13.9%	94.0%	42
ESI*	**1254**	**875**	**0.8%**	**82.9%**	**37**
S14-1941	1714	1068	0.9%	76.0%	45
S14-1944	757	626	0.5%	88.4%	30
S14-1962	1581	1066	0.8%	81.3%	40
S14-2196	963	740	1.1%	85.9%	32
LDI*	**4111**	**2664**	**19.8%**	**75.4%**	**82**
S14-1941	3696	2678	22.6%	82.6%	78
S14-1944	4626	2604	18.6%	66.1%	86
S14-1962	4254	2866	19.1%	77.3%	80
S14-2196	3869	2506	18.9%	75.4%	83
Grand Total**	**2264**	**1596**	**15.4%**	**85.9%**	**54**

* Average values across 4 samples; **** Average values across all samples and ionisation techniques

Errors of Assignments – Histograms

Figure S1 shows the distribution of accurate mass errors as histograms for each ionisation mode across the cumulative assignment of formula for the four whisky samples. Formula were assigned with a ±0.25 ppm error threshold, reflected in the limits of the plots. A narrow distribution around 0 ppm error is typical for well calibrated and accurately assigned formulas.[8] Confidence decreases as the distribution spreads or shifts. Plots produced using *Seaborn 0.8.1* and *Matplotlib 2.1.1* on *Python 3.6.4*.

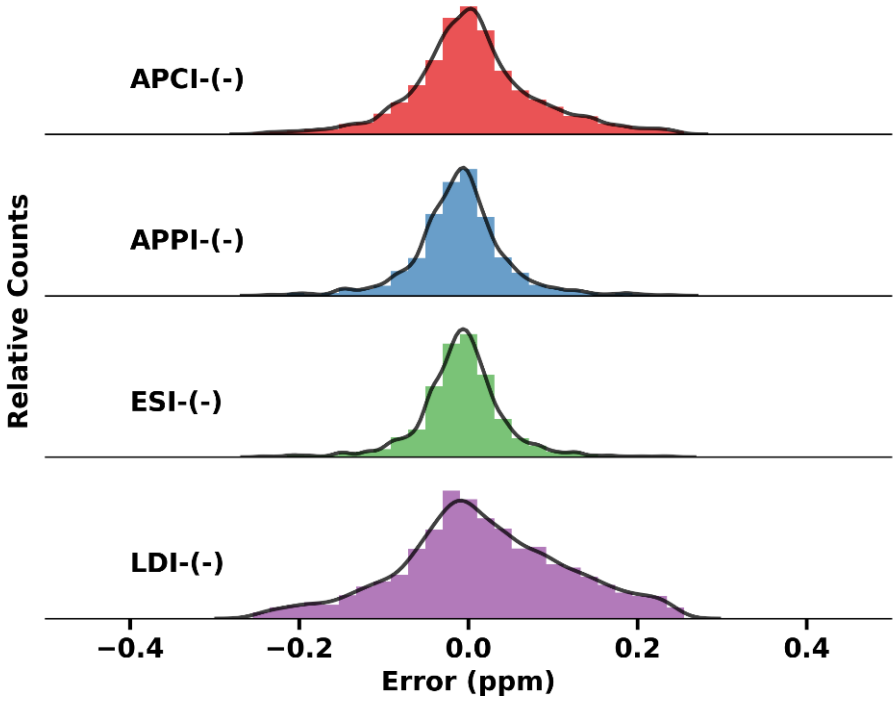

Figure S1 - Histogram for the errors of assignments for each ionisation source. Y-Axes are scaled independently for each histogram.

Radical Anion Assignment

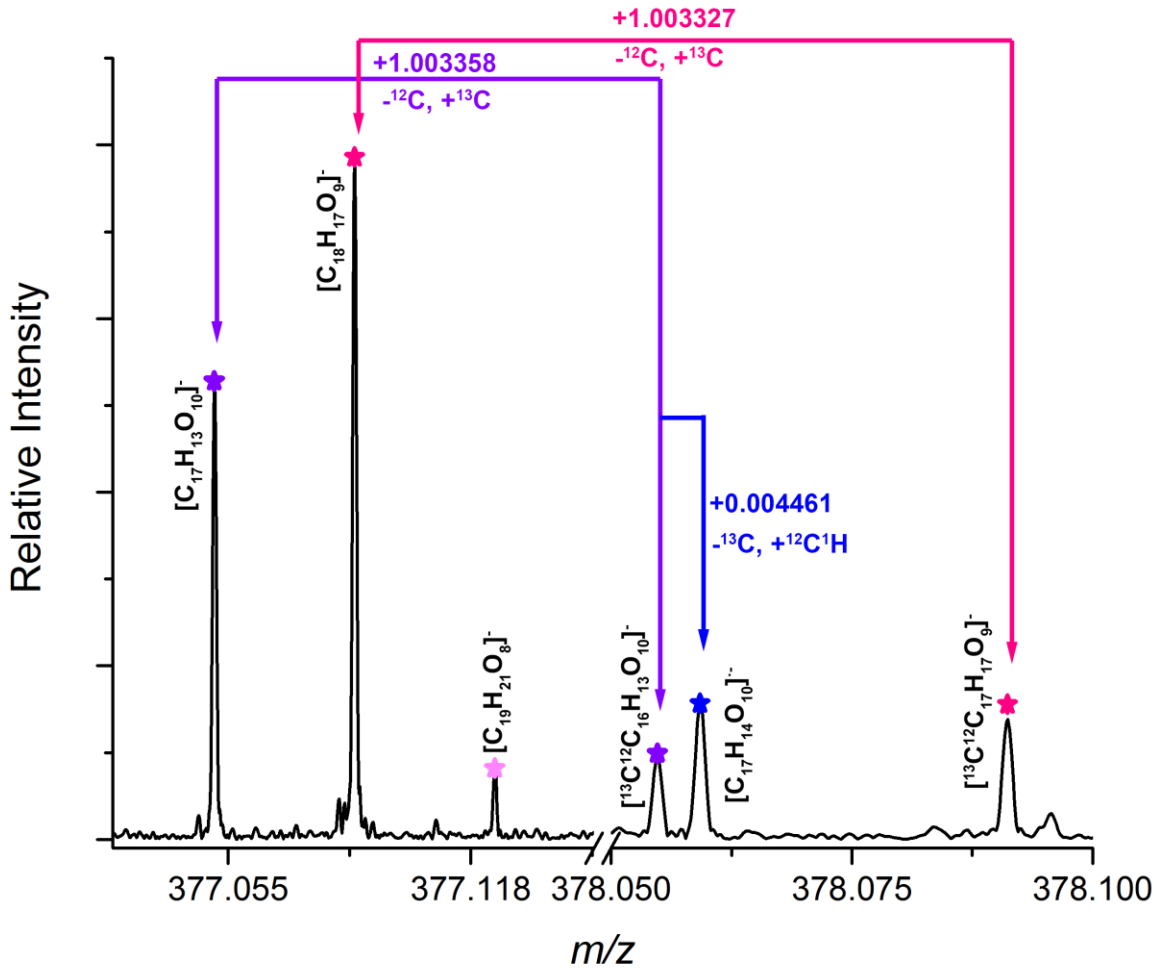

Figure S2 - Region of APPI (-) mass spectrum for sample S14-1941 at m/z 377-378 showing several assignments, including one compound ($C_{17}H_{14}O_{10}$) ionised both as deprotonated and radical anion form. Mass differences have been annotated to aid inspection. Theoretical peak positions and intensities are plotted as coloured scatter marks with star glyphs.

Table S4 – Summary of molecular formula assignments as radical anions as an average across 4 samples

	New Formula from Radicals*	**Deduplicated Formulas****	**% New Formula from Radicals*****
APCI	91	1483	6.1%
APPI	2	686	0.3%
ESI	3	871	0.3%
LDI	27	2163	1.3%
Grand Total Averages	**31**	**1301**	**2.4%**

***Number of formula assigned as radicals which were not identified from the deprotonated ion form.

**The average total number of formula after de-duplication of radicals and deprotonated species

***Percentage of radical only formula relative to deduplicated formula

Formula Assignment Statistics

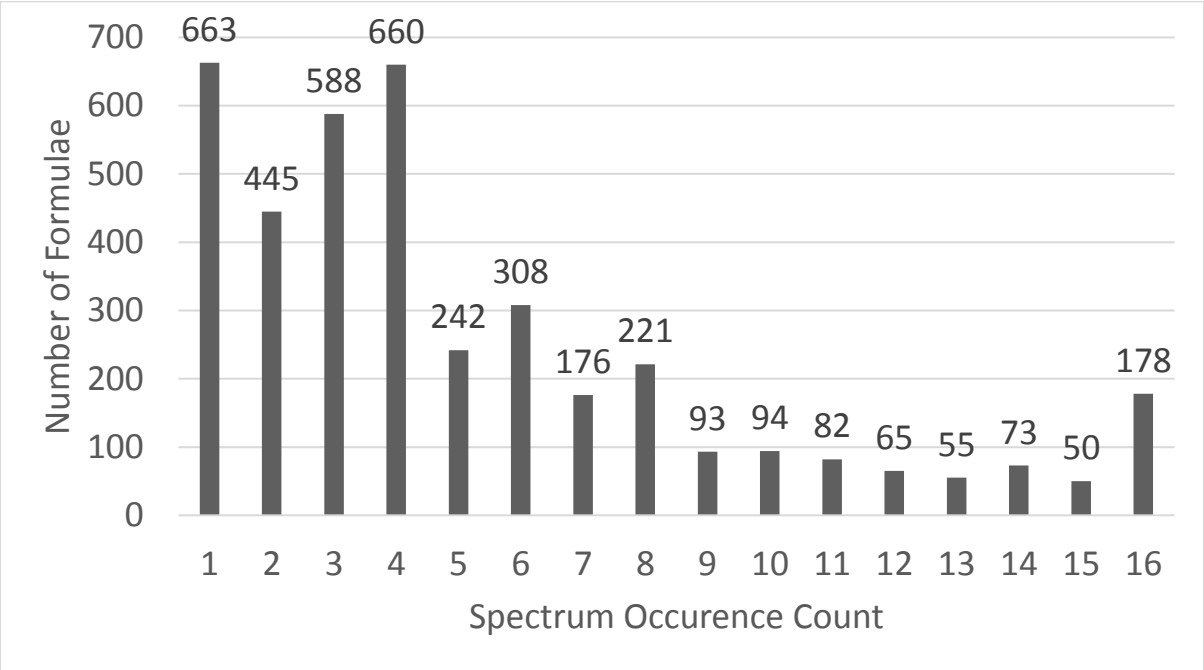

Figure S3 – Number of times each unique formula (n=3993) appears across the 16 acquired spectra. Figure created using Microsoft Excel 2016 PivotTable functionality.

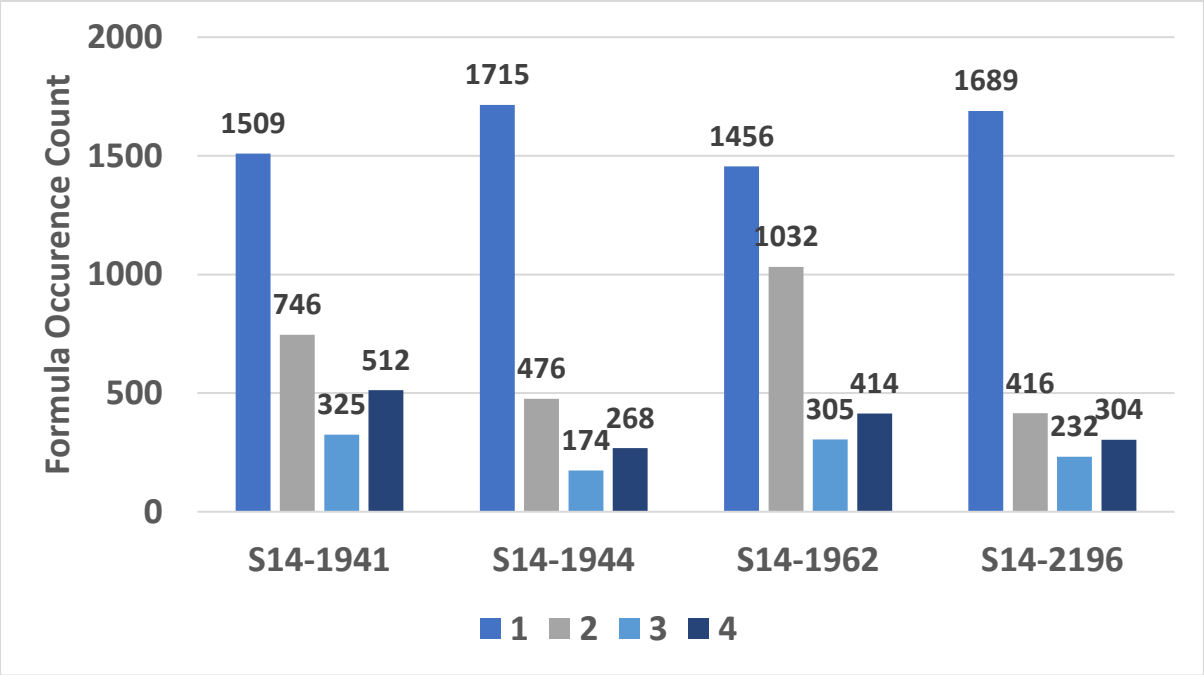

Figure S4 - Number of times each unique formula appears across four ionisation modes for individual whisky samples.

Formula Common to All, van Krevelen diagram

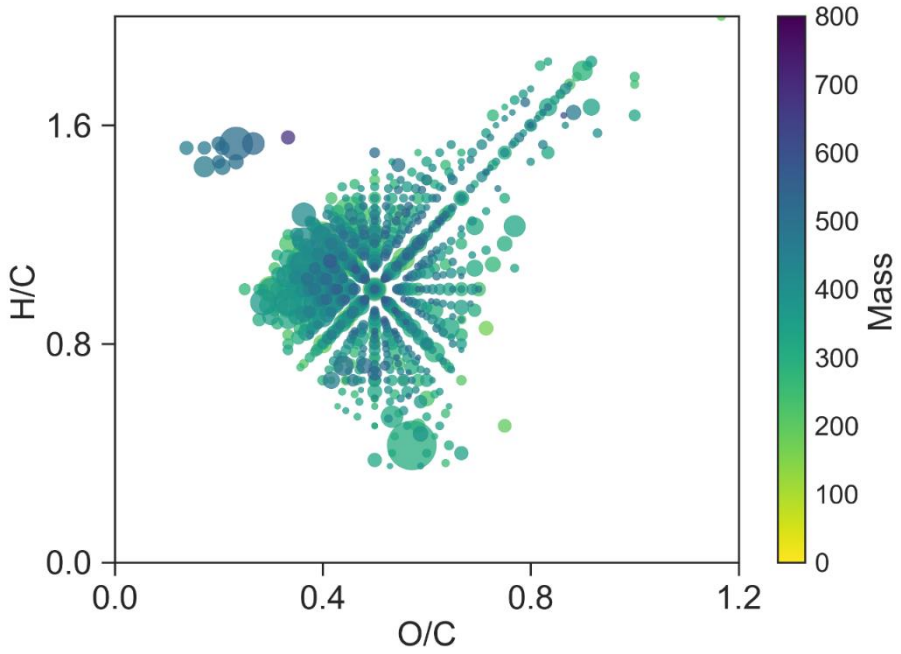

Figure S5 - Van Krevelen diagram for 699 formulas common to all four ionisation techniques.

Unique and Common Formula, van Krevelen diagram

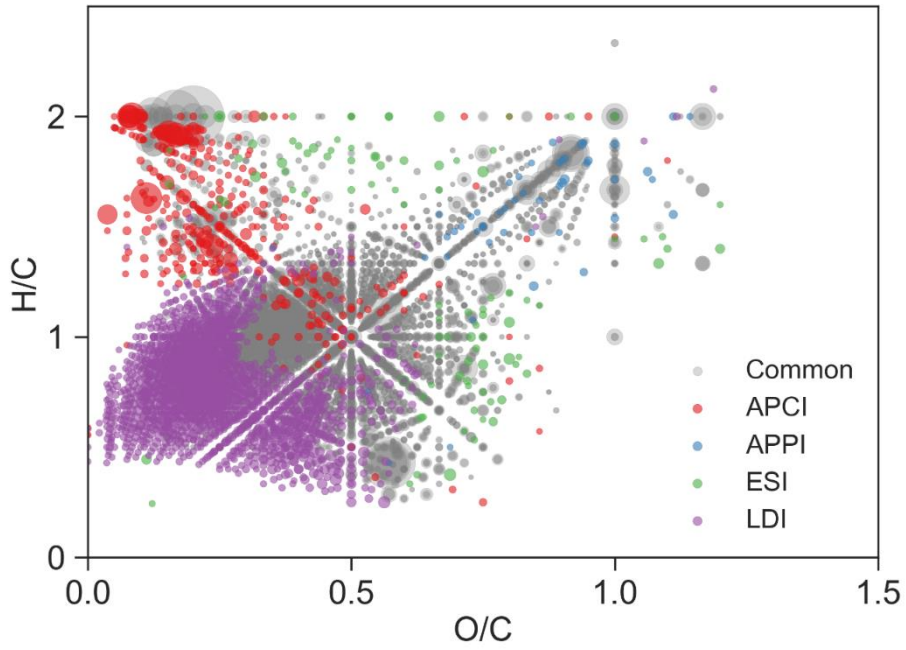

Figure S6 – Van Krevelen plots showing formula unique to a given ionisation source. Glyphs are colour coded by their ionisation source, and size coded relative to median intensity of each given formula across multiple samples. Grey glyphs represent formula not unique to a given ionisation source and are 'common' to two or more sources. Grey glyphs are plotted first and can be obscured by overlapping points representing the other ionisation techniques.

Unique and Common Formula, DBE vs C plots

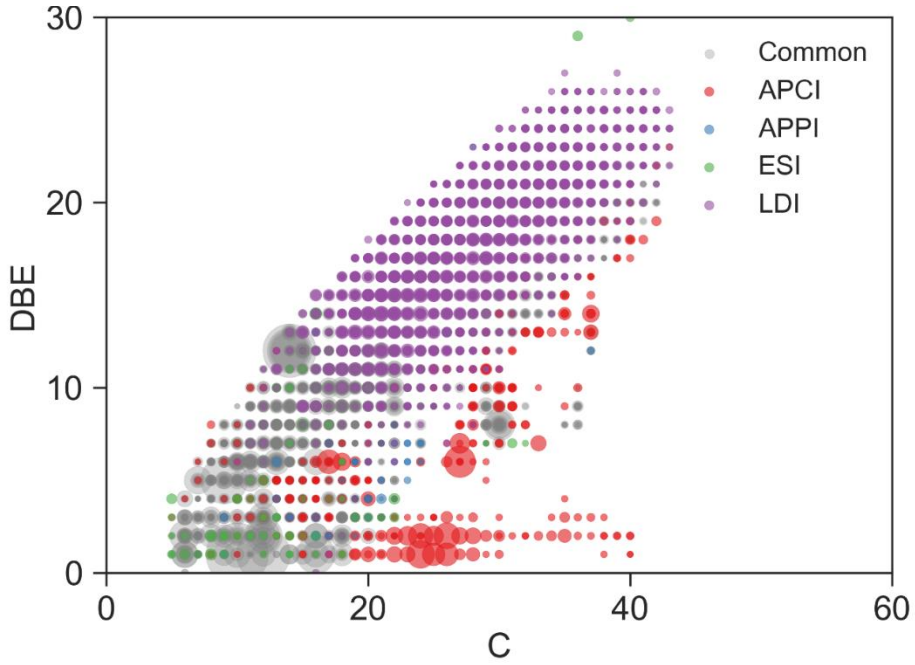

Figure S7 - DBE v Carbon number plots showing formula unique to a given ionisation source. Glyphs are colour coded by their ionisation source, and size coded relative to median intensity of each given formula across multiple samples. Grey glyphs represent formula not unique to a given ionisation source and are 'common' to two or more sources. Grey glyphs are plotted first and can be obscured by overlapping points representing the other ionisation techniques.

Aromaticity Index Bar Plot

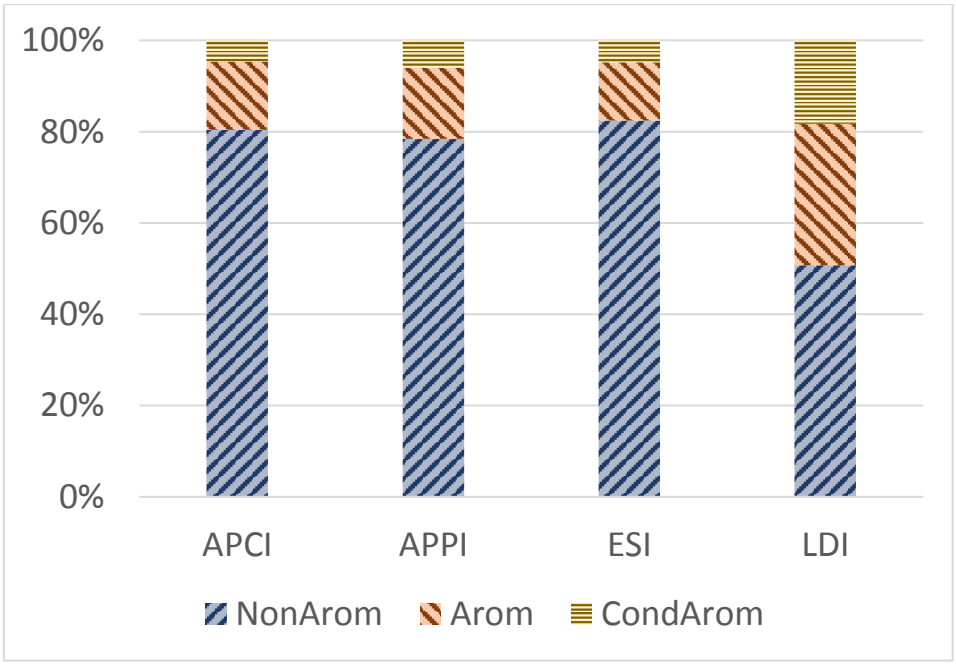

Figure S8 – Modified Aromaticity Index (AI_{mod}) bar plot. Counts of formula have been normalised to 100% individually for each ionisation technique to compensate for significantly more formulas identified in LDI and APCI than APPI and ESI. Aromaticity Index calculated as per Koch et al.[9] The modified AI is calculated with the number of oxygens halved. This estimates the number of sigma-bound oxygen atoms and produces a more realistic assessment of the aromaticity index

References

(1) Kew, W.; Blackburn, J. W. T.; Clarke, D. J.; Uhrín, D. *Rapid Commun. Mass Spectrom.* **2017**, *31* (7), 658–662.

(2) Cho, Y.; Ahmed, A.; Islam, A.; Kim, S. *Mass Spectrom. Rev.* **2015**, *34* (2), 248–263.

(3) Cho, E.; Witt, M.; Hur, M.; Jung, M. J.; Kim, S. *Anal. Chem.* **2017**, *89* (22), 12101–12107.

(4) Schweihard, L.; Lindinger, M.; Kluge, H.-J. *Int. J. Mass Spectrom. Ion Process.* **1990**, *98* (1), 25–33.

(5) Schweikhard, L. *Int. J. Mass Spectrom. Ion Process.* **1991**, *107* (2), 281–292.

(6) Tolić, N.; Liu, Y.; Liyu, A.; Shen, Y.; Tfaily, M. M.; Kujawinski, E. B.; Longnecker, K.; Kuo, L.-J.; Robinson, E. W.; Paša-Tolić, L.; Hess, N. J. *Anal. Chem.* **2017**, *89* (23), 12659–12665.

(7) Kujawinski, E. B.; Behn, M. D. *Anal. Chem.* **2006**, *78* (13), 4363–4373.

(8) Smith, D. F.; Kharchenko, A.; Konijnenburg, M.; Klinkert, I.; Paša-Tolić, L.; Heeren, R. M. A. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (11), 1865–1872.

(9) Koch, B. P.; Dittmar, T. *Rapid Commun. Mass Spectrom.* **2006**, *20* (5), 926–932.